



Angus McIntosh Centre
for Historical Linguistics

LEVERHULME
TRUST

El *Sermón en Lengua de Chile* en el *Corpus Histórico del Mapudungun*

De texto colonial a texto digital

Benjamín Molineaux

26 de agosto 2021

Simposio conmemorativo

“400 años de *Sermón en Lengua de Chile*”



Corpus Histórico del Mapudungun
Corpus of Historical Mapudungun



THE UNIVERSITY
of EDINBURGH

Objetivos

- ◉ Introducción al *Corpus Histórico del Mapudungun* (*Trokiñlechi Chilka Kuyfike Mapudungun*)
- ◉ El *Sermón en Lengua de Chile* en el CHM
 - Compilación
 - Estructura
- ◉ Posibles usos del corpus
- ◉ Invitación a usarlo



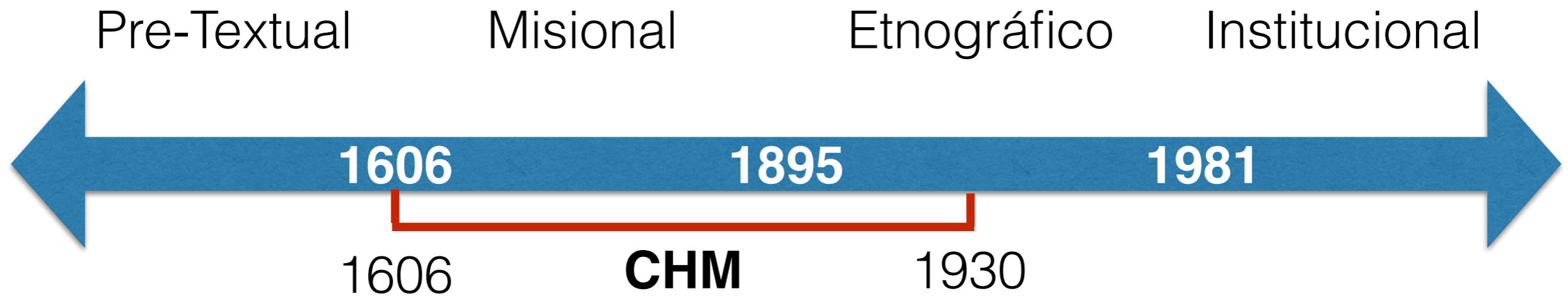
El Corpus Histórico del Mapudungun

- Un proyecto postdoctoral (Leverhulme) de tres años
 - En el *Angus McIntosh Centre for Historical Linguistics*
 - *Objetivo:* Producir un corpus de materiales en mapudungun histórico que permita búsquedas a nivel
 - Léxico
 - Morfológico
 - Fonológico
 - De metadatos (año, lugar, autoría, etc.)
- Versión 1.0 disponible y de libre acceso



Lingüística Histórica Mapuche

- A pesar de 400 años de materiales, el área se ha explorado poco
- A menudo se tratan distintos dialectos y períodos como uno sólo
- Los materiales no son de fácil acceso
- La calidad de los datos empeora progresivamente al retroceder



(Villena, 2017)



Los textos del Corpus

<https://benmolineaux.github.io/bookshelf/sp/>

Año	Título	Compilador	Orador(a)
1606	Doctrina y Catequismo	Valdivia	?
1621	Sermón en lengua de Chile	Valdivia	?
1642	Vocabula Chilensia	Herkmans?	?
1765	Koyagtun, Diálogo, Doctrina	Febrés	?
1777	Índice Universal, Catequismo y Vocabulario	Havestadt	?
1879	Fraseario	Barbará	Felipe Marinao ?
1897	Estudios Araucanos	Lenz	Kalfün et al.
1903	Nidolke dəŋu Dios ñi Nùtram	Augusta	?
1903	Gramatica Araucana (Lecturas)	Augusta	?
1910	Lecturas Araucanas	Augusta/Frauenhäusel	Painemilla, Wenuñamko, et al
1922	Pismahuile: Un cuento araucano	Augusta/Frauenhäusel	Wenuñamko
1913	Últimas Familias	Guevara	Manquilef
1911	Folklore Araucano	Guevara	Manquilef
1911	Comentarios al Pueblo Araucano (Faz Social)	Manquilef	Manquilef
1930	Vida y costumbres ... (Pascual Coña)	Mösbach	Pascual Coña



El Sermón en Lengua de Chile en el CHM

- Es una obra extensa, temprana y muestra un conocimiento avanzado de la lengua

“La precision y claridad expositiva de Valdivia es tal, que estamos por asegurar que se trata de uno de los gramáticos que mejor ha explicado estas cuestiones en el siglo XVII y quien mas consecuentemente ha decidido practicar una ortografía coherente de las lenguas indígenas.”

Calvo, 2005



Generando el texto digital

- Desde imágenes digitales, se creó un texto de lectura electrónica
 - Optical Character Recognition (OCR)
 - Corrección manual de errores de OCR

SERMON SEGVNDO
DE LA GRAVEDAD DEL PE
cado, por lo mucho que a Dios ofende
y al hombre daña, y por los castigos
que Dios ha hecho en este mun-
do por pecados. 81

¶ Si acaso os acordays
(hermanos) en el sermon
passado os vine a dezir como
ay otra vida despues de
esta donde nuestra anima
viue para siempre. ¶ Y q̄
aquel gran Señor Dios, ha-
zedor de todas las cosas, a
los buenos que le siruen les
da bienes de gozo, y de esc̄a-
so eterno, porque quiere
mucho al hombre, y lo crió
a su semejança, y tiene co-
mo a hijo. Y assi tiene guar-
dado para los que son
buenos

T Vme chey duamtuyrn, (pu peñi ema)
ca sermon meu, ta ni gen ta ca mogen,
able vachi mogen, ta piuyeyrn ta cū
paūyen. Veymeu ta in pu pllu mo-
gey ta mgenque. Teye vita apo Dios vill du-
gu vemvoe, ta ni uvchiqueeteu cume que pu
che meu cūme repen dugu abnoluchi urcūn cay ra
eluquebi; ta ni aldu ayubiyum ta che; quidu ta ni gen
mgelu

9SermonesVisionMapu — Edited
SERMON SEGVNDO
DE LA GRAVEDAD DEL PE
cado, por lo mucho que a Dios ofende
y al hombre daña, y por los castigos .
que Dios ha hecho en este mun-
do por pecados. SI
222 Vme chey duamtuyrn, (pu peñi ema)
ca sermon meu, ta ni gen ta ca mogen,
s able vachi mogen, ta piuyeyrn ta cū
Sense paūyen. Veymeu ta in pu pllu mo-
gey ta mgenque. Teye vita apo Dios vill du-
gu vemvoe, ta ni uvchiqueeteu cume que pu
che meu cūme repen dugu abnoluchi urcūn cay ra
eluquebizontani aldu ayubiyum ca che; quidutani gen
EA
Es
mgelu

- Agradezco a David Fenton-Smith y Patricia Ojeda

Lematización

- El objetivo es proveer una etiqueta única para palabras relacionadas, eliminando heterogeneidad a nivel morfológico y ortográfico
 - Conversión automática al Alfabeto Mapuche Unificado (Catrileo)
 - Cada palabra en AMU se asocia un lema
lema = palabra – flexión (+ (ü)n para los verbos finitos)
 - Se asocia una categoría léxica (POS) a cada lema
- Este lema se extiende a cada instancia de la palabra en el corpus

Ortografía	AMU	Lema	POS	Fuente	Año
<hueda-layay>	wedalayay	wedan ‘ser/estar mal/o/a’	V	Sermones	1621
<hueda-lay>	wedalay	wedan ‘ser/estar mal/o/a’	V	Sermones	1621
<we r a-y>	weday	wedan ‘ser/estar mal/o/a’	V	Pismahuile	1921

- Codificado en formato XML, usando estándares TEI:

```
<w Lemma="wedan" pos="V" corresp="be bad" sameAs="ser malo">huedalay</w>
```



Anotación morfológica

- Provee una manera de identificar y comparar las raíces y sufijos
- Consiste en identificar los morfemas de cada palabra y asignarles una etiqueta
 - El proceso se expande a todas las palabras de estructura similar

Ortografía	Morfemas	Lema	POS
<lag-ùm-a-bi-mi>	<i>morir</i> -CAUS-FUT-3DIR-IND.2.S	l'angümün 'matar'	V
<·La-uye-y>	<i>morir</i> -PERF-IND.3	l'an 'morir'	V

```
<w xml:lang="arn" lemmaRef="l'auyey" pos="V" corresp="die" sameAs="morir">  
  <m baseForm="l'a" type="root" corresp="die" sameAs="morir">·La</m>  
  <m baseForm="uye" type="sfx" subtype="inflectional" corresp="perf" sameAs="perf">uye</m>  
  <m baseForm="iy" type="sfx" subtype="inflectional" corresp="ind3" sameAs="ind3">y</m>  
</w>
```



Anotación grafo-fónica

- El mapeo de la ortografía (grafemas) al sistema de sonidos (fonos)
- Este es un proceso semi-automático, basado en las indicaciones de los mismos compiladores (en este caso, Valdivia 1606) y traducido al Alfabeto Fonético Internacional (AFI)
- Permite realizar comparaciones a lo largo del espacio y del tiempo

Ortografía	Sonidos	Lema	Dialecto	Fuente
<vùta>	[βwta]	fücha 'grande/viejo'	Centro/Norte	Sermones, 1621
<fücha>	[fɪtʃa]	fücha 'grande/viejo'	Panguipulli	Pismahuile, 1921

```
<w lemma="fücha" pos="AJ" corresp="old/big" sameAs="grande/viejo">
  <m baseForm="fücha" type="root" corresp="old/big" sameAs="grande/viejo">
    <c corresp="β">v</c>
    <c corresp="w">ù</c>
    <c corresp="t">t</c>
    <c corresp="a">a</c>
  </m>
</w>
```





Angus McIntosh Centre
for Historical Linguistics

Usos del *Corpus Histórico del Mapudungun*



Corpus Histórico del Mapudungun
Corpus of Historical Mapudungun



THE UNIVERSITY
of EDINBURGH

Uso 1: Acceso directo a los textos

- Permite la lectura, copia y reutilización de los textos digitales
- En la superficie, se presenta la ortografía original
- La lematización y anotación morfológica está disponible visualmente

Sermon Segundo: De La Gravedad del Pecado, por lo mucho que a Dios ofende y al hombre daña, y por los castigos . que Dios ha hecho en este mundo por pecados.

LEMMA: ngelyüm

POS: VNF

ENG: while creating

SPA: al crear

MORPH: nge-(ü)l-fi-

yüm

be-CAUSE-3DIR-TEMP

(ser/haber)

1

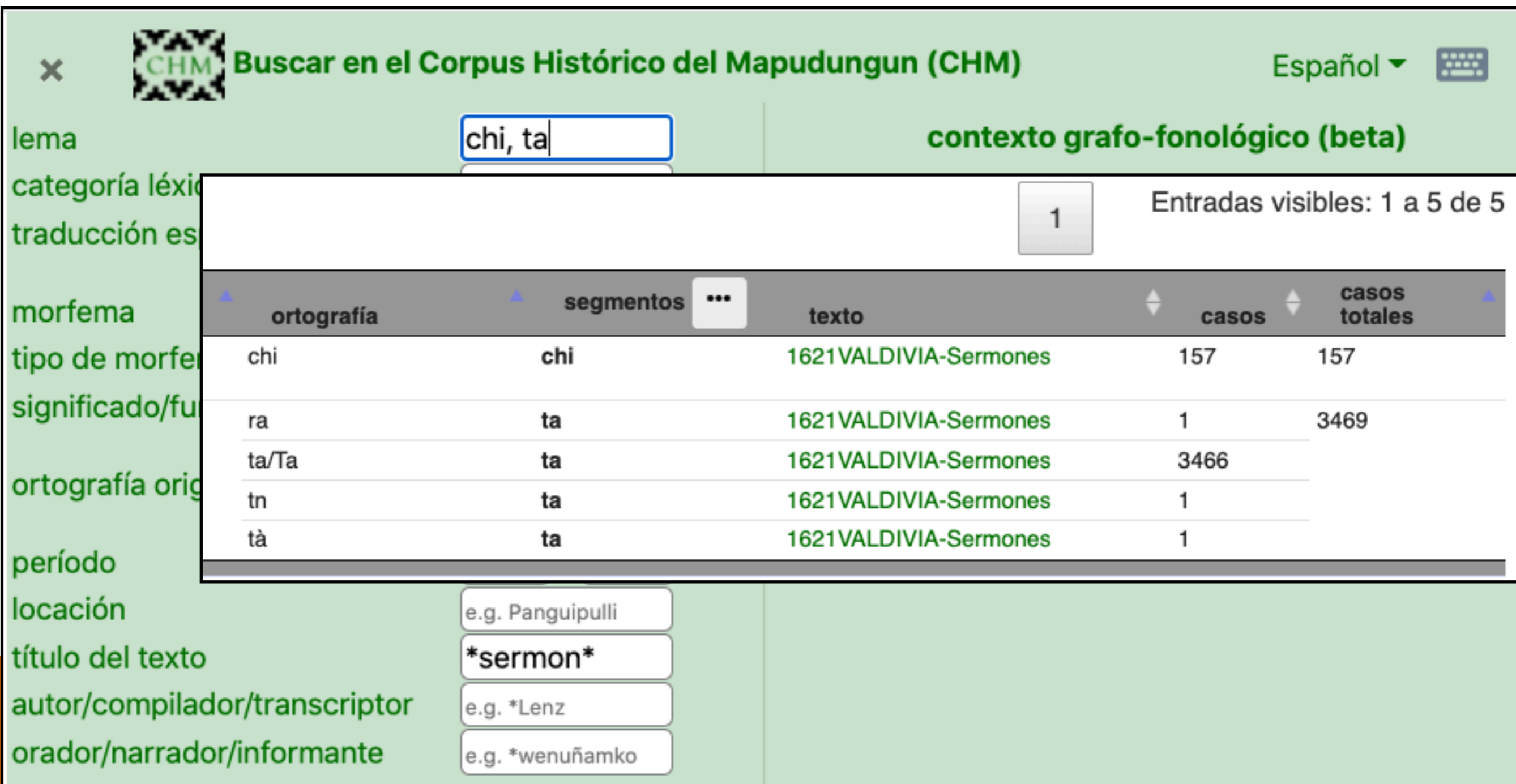
mn, (pu peñi ema) ca sermon meu,
ble vachi mogen, ta piuyeeymn ta
pu pllù mogey ta mgenque. Teye
vemvoe, ta ñi uvchiqueeteu cùme
pendugu abnoluchi ùrcùtun cay ra
iyùm ta che; quidu tañi gen mgelu
ta ñi gelbiyùm ta che cay ta Dios, votùm yebi cay.
Vamgechi cay ta pu che meu vey ta ñi cùme votùm gelu, ta
ñi piel chi dugu mùpituquelu pu ve, vùta que dugu cùme
gelu mgenquenmayalu nieelcanmabi: huall quiñe queun
rume pepi raquinoalu, quiñe duam rume pepi duamtunoalu
vachi cùme que dugu.

1 ¶ Si acaso os acordays (hermanos) en el sermon pasado os vine a dezir como ay otra vida despues de esta donde nuestra anima viue para siempre. ¶ Y que aquel gran Señor Dios, hazedor de todas las cosas, a los buenos que le siruen les da bienes de gozo, y descanso eterno, porque quiere mucho al hombre y lo crio a su semejança, y tiene como a hijo. Y assi tiene guardados para los que son buenos hijos, y guardan sus mandamientos bienes eternos tan grandes que no ay lengua que pueda contarlos, ni coraçon que pueda pensallos.



Uso 2: Búsqueda por etiquetas (tablas)

- Es un formulario web hecho a medida
 - Permite hacer búsquedas de las etiquetas XML
 - Presenta los resultados como tablas con hipervínculos



CHM Buscar en el Corpus Histórico del Mapudungun (CHM) Español

lema contexto grafo-fonológico (beta)

1 Entradas visibles: 1 a 5 de 5

ortografía	segmentos	texto	casos	casos totales
chi	chi	1621VALDIVIA-Sermones	157	157
ra	ta	1621VALDIVIA-Sermones	1	3469
ta/Ta	ta	1621VALDIVIA-Sermones	3466	
tn	ta	1621VALDIVIA-Sermones	1	
tà	ta	1621VALDIVIA-Sermones	1	

e.g. Panguipulli

sermon

e.g. *Lenz

e.g. *wenuñamko

lema

categoria léxica

traducción española

morfema

tipo de morfema

significado/función

ortografía original

período

locación



título del texto

autor/compilador/transcriptor

orador/narrador/informante


Uso 3: Búsqueda grafo-fónica (Medusa)


- Es un buscador visual
 - Permite interacción directa con los grafemas y fonos
 - Presenta hipervínculos a los resultados en tablas

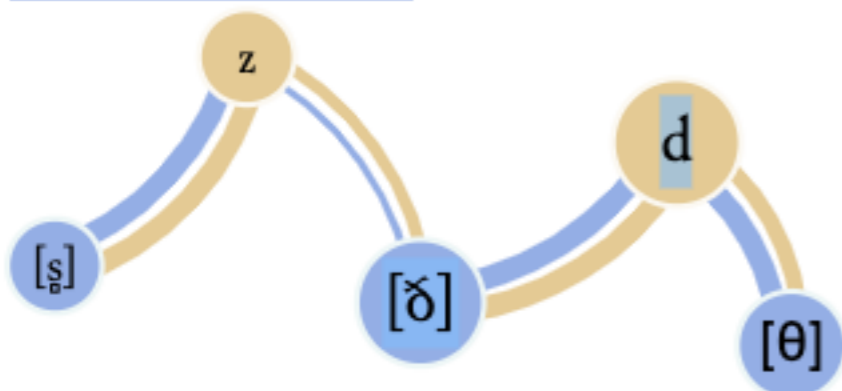
 **Buscar en el Corpus Histórico del Mapudungun (CHM)** Español 

lema
categoría léxica
traducción española

ortografía
fono

<z> 
casos totales: 30
<z> -[s] : 25
<z> -[ð] : 5

<d> 
casos totales: 1770
<d> -[ð] : 1606
<d> -[θ] : 164



<d>-[ð]
morfemas: 68, ortografías: 280, casos: 1605, textos: 1

Por morfema

Por texto

texto	año
1621VALDIVIA-Sermones	1621
1922-Pishmawile	1922



Aplicaciones del CHM

- **Académicas:**

- Búsquedas de patrones lingüísticos a nivel
 - Léxico
 - Morfológico
 - Fonológico
- Util en otras disciplinas
 - Historia, antropología, estudios culturales, teología, literatura, etc.

- **Non-Académicas:**

- Testimonio del patrimonio cultural mapuche
- Material para la revitalización/preservación de la lengua
 - Un recurso para los hablantes y aprendices de la lengua
 - Una fuente de estructuras para la planificación lingüística





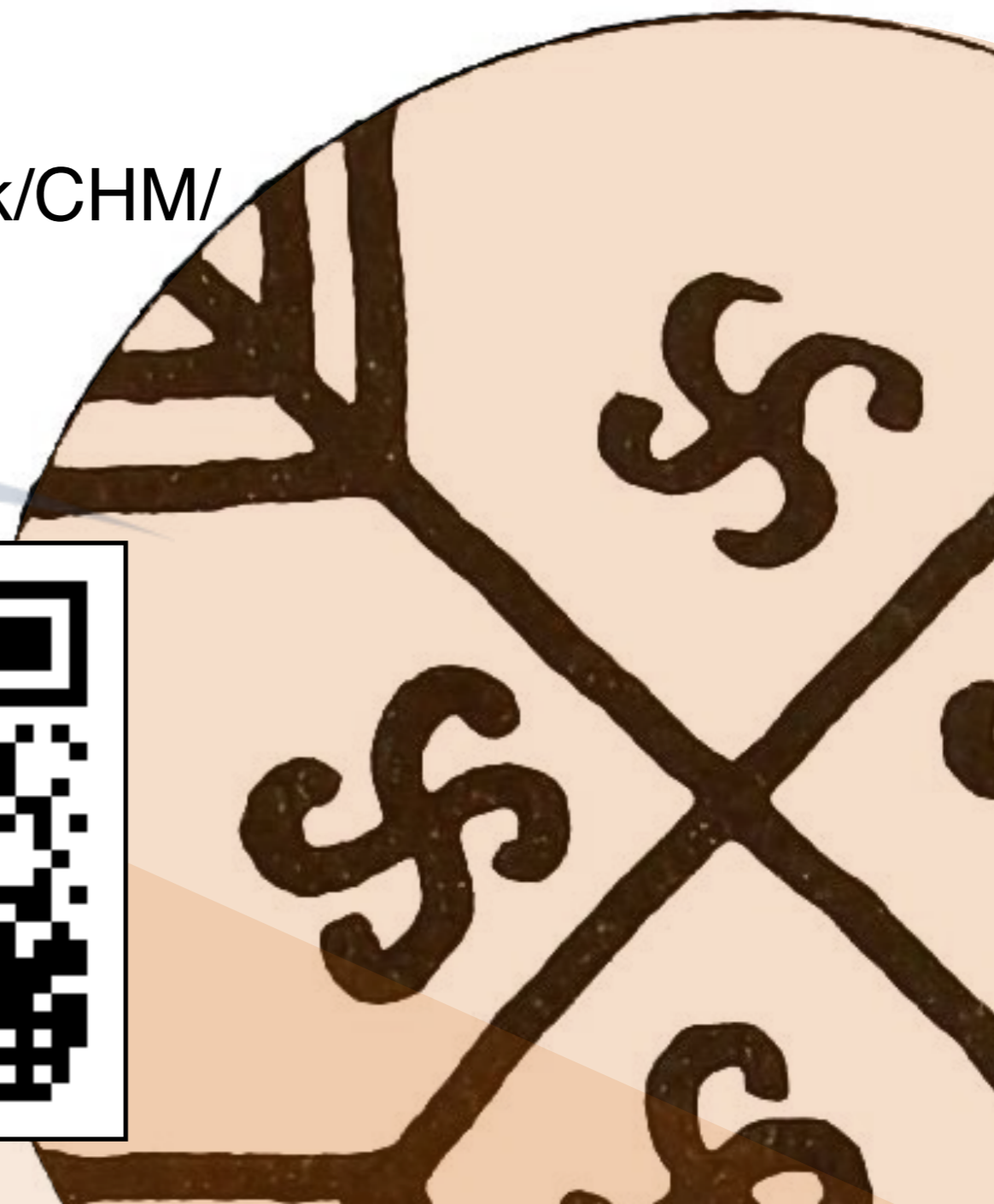
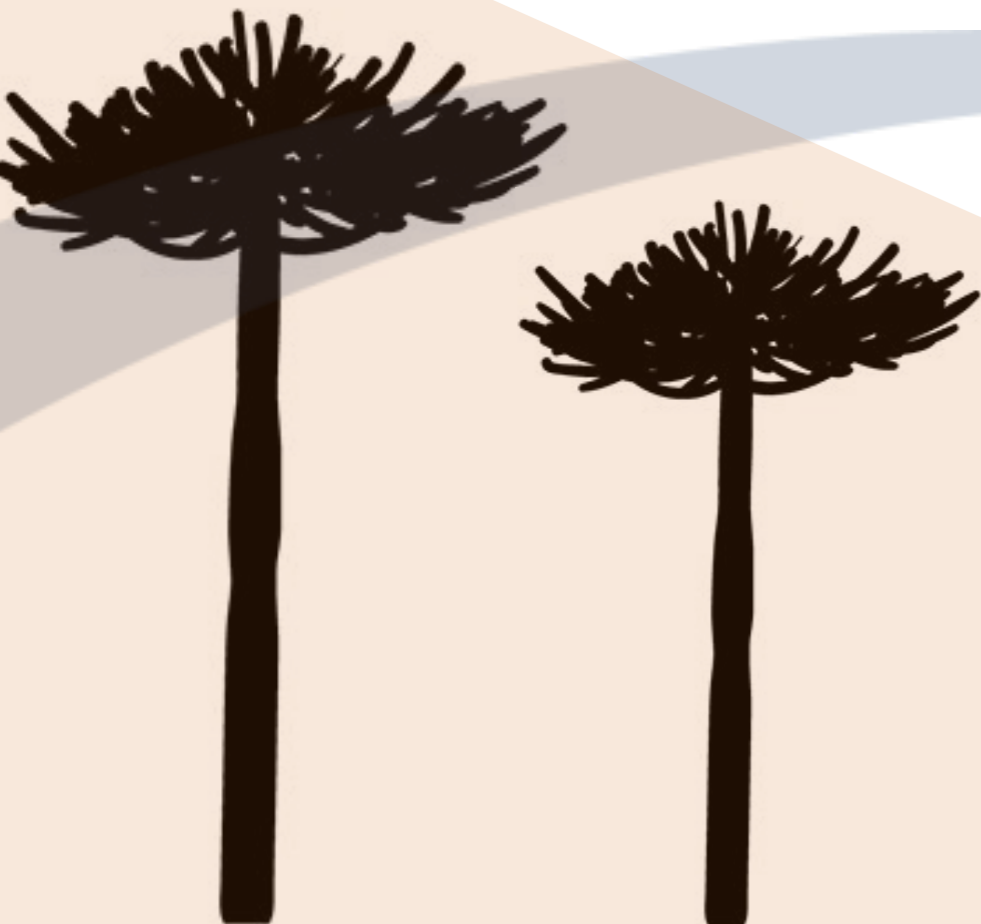
Angus McIntosh Centre
for Historical Linguistics



THE UNIVERSITY
of EDINBURGH

¡Chaltu may!

<http://www.amc-resources.lel.ed.ac.uk/CHM/>






Bibliografía selecta

- Calvo, J. 2005. Fonología y ortografía de las lenguas indígenas de América del Sur a la luz de los primeros misioneros gramáticos. En Zwartjes, O. y Altman C. (eds.) *Missionary Linguistics II*. Amsterdam/Philadelphia: John Benjamins: 137–170.
- Catrileo, M. 1986. La variación estilística en el nivel fonológico del mapudungun. *Actas de Lengua y Literatura Mapuche 2*, 1-19.
- Chandía, A., B. Villena, Araya I., Flores, C. and Leiva, C. 2021. Corpus lexicográfico del mapudungun (CORLEXIM). <http://www.chandia.net/corlexim>.
- Croese, R., A. Salas, and G. Sepúlveda. 1978. Proposición de un sistema unificado de transcripción fonémica para el mapuche. *Revista de Lingüística Teórica y Aplicada*, 16, 151-159.
- Molineaux, B. 2021. Source Material for the Corpus of Historical Mapudungun. © The University of Edinburgh, URL: <https://benmolineaux.github.io/bookshelf/>.
- Molineaux, B. and V. Karaiskos (2021) Corpus of Historical Mapudungun (Version 1.0), © The University of Edinburgh, URL: <http://www.amc-resources.lel.ed.ac.uk/CHM/>.
- Tarpley, B. 2018. Digital humanities dashboard. Center of Digital Humanities Research, Texas A&M University. <http://codhr.dh.tamu.edu/2018/04/24/the-early-modern-ocr-project/>.
- TEI Consortium, eds. 2021 Guidelines for Electronic Text Encoding and Interchange v.4.2.2. 09 April 2021. <http://www.tei-c.org/P5/>.
- Villena, B. 2017. Fuentes para el estudio del mapudungún: propuesta de periodización. *Lenguas y literaturas indoamericanas*, 19(1), pp. 141-167.

Investigaciones Léxicas (SLCh)

- (falta de) prestamos

  **Buscar en el Corpus Histórico del Mapudungun (CHM)**

Español ▾ 

lema

categoría léxica

traducción española

lema	categoría	morfemas	ortografía
ifum	N	ifum oveja	ybum yvum
weke	N	weke oveja/camelido pequeño	hueque



Investigaciones Morfológicas (SLCh)

- **Sufijos poco estudiados**

che # dke persona # ALONE	Chedque/chedque
Cristo # dke Cristo # ALONE	Christodque
diaflo # dke diablo # ALONE	diablodque

....., vill ta abi ta che; pura **chedque** montuy,
.....
se acabaron todos, solo se escaparon ocho personas, d

-dke = 'tan solo'; 'alone'



Investigaciones Fonológicas (SLCh)

Variación estilística

- Hoy: [ʃ/z]~[θ/ð]~[ʃ]~[s]
(Catrileo 1986)

mürke, müdke, müshke (harina tostada)
dewman, sewman, shewman (hacer)

weda # la # iy malo # NEG # IND3	huedalay
weda # nge # chi malo # VBLZ # AJVZ	hueragechi
weda # nge # lu malo # VBLZ # SVN	hueragelu
weda # nge # fu # iy malo # VBLZ # BI # IND3	hueragebuy

pod polvo/suciedad	por
pod # el # nge # tu # (ü)n polvo/suciedad # NMLZ # PASS # REST # INF	porgetun
pod # nge # tu # iy polvo/suciedad # VBLZ # TR # IND3	porgetuy
pod # nge # tu # (ü)n polvo/suciedad # VBLZ # REST # INF	porgetun

